Buletin Geoinformasi, Jld. 3 No. 1, ms. 13 - 21, Sept, 1999
I Penerbitan Akademik Fakulti Kejuruteraan & Sains Geoinformasi

PELARASAN L! NORM DAN PENGESANAN SELISIH KASAR CERAPAN UKUR

Halim Setan, PhD. Yul
Amri Muhamad Yunus
Center for Industrial Measurement and Engineering Surveying (CIMES)
Fakulti Kejuruteraan dan Sains Geoinformasi
Universiti Teknologi Malaysia
e-mail:-
Halim(fl>.fksa.utm.mv

Abstrak

Kaedah pelarasan ukur dan ujian statistik yang digunakan akan menentukan keberkesanan pengesanan selisih kasar dalam satu jaringan geodetik. Berasaskan kepada pelarasan kuasa dua terkecil (L_2 - Norm), beberapa kaedah bagi mengesan selisih kasar telah dikembangkan, seperti kaedah Baarda, Pope (Tau) dan Danish. Berbasis kepada pelarasan L_2, kegagalan pengesanan mungkin akan semakin meningkat apabila wujud selisih kasar berganda, meningkatnya korelasi reja dan rendahnya bilangan lebihpada. Selain daripada L_2- Norm, satu lagi kaedah pelarasan menggunakan kriteria jumlah absolut reja terkecil (Li-Norm). Berbeza dengan pelarasan L_2 yang sensitif terhadap selisih kasar, pelarasan L, tidak sensitif terhadap selisih kasar. Kertas kerja ini membincangkan pelarasan dan prosedur pengesanan selisih kasar berdasarkan LI. Satu contoh numerikal disertakan untuk menunjukkan keberkesanan kaedah pengesanan tersebut.

1.0 PENGENALAN

I '.eh sebab cerapan ukur selalunya mengandungi kesalahan (iaitu selisih kasar), maka untuk -enentukan n parameter selalunya dilakukan m cerapan, dimana m > n. Persoalannya kemudian iiaiah wujud m!/((m-n)!.n!) kombinasi daripada persamaan cerapan-parameter yang boleh ~emberikan penyelesaian yang berbeza-beza. Satu penyelesaian tunggal boleh diperolehi dengan ~e:r.perkenalkan satu kriteria terhadap reja cerapan. Tiga kriteria yang biasa digunakan ialah kriteria _-lah mutlak terkecil (disebut LI), kuasa dua terkecil (disebut L_2) dan minima-maksima (L_m).

Serarang kriteria boleh dipilih. Namun, justifikasi teoritikal dalam pemilihan kriteria bermula :z:;pada pengetahuan tentang model stokastik bagi cerapan. Pelarasan L_2 berasaskan kepada bahawa selisih rawak cerapan (dan reja v) mengikut taburan kebarangkalian normal. ini mula tergugat apabila satu set data cerapan daripada 21,365 reja menunjukkan bahawa --r-ran kebarangkalian daripada selisih rawak lebih hampir kepada taburan g(v) = 1/2 . h . e" " M -;-randing taburan normal (Branham, 1990), dimana h adalah ukuran daripada ketumpuan. Apabila .- :i: stokastik daripada selisih mengikut taburan di atas, maka pelarasan L_t dikatakan lebih realistik :ar:?ada L_2.

-^ arasan L_2 sangat peka terhadap wujudnya selisih kasar (Halim, 1995) manakala L, tidak Barrodale & Roberts, 1974; Branham, 1990). Ini bermakna bahawa hasil pelarasan L, secara -e r.ifnya tidak terpengaruh oleh wujudnya selisih kasar dan oleh itu boleh digunakan terus untuk -e-.Liesan selisih kasar (Marshall & Bethel, 1996b; Fuchs, 1982).

kepada pelarasan kuasa dua terkecil terdapat sekurang-kurangnya 6 kaedah bagi _esan selisih kasar. Empat kaedah yang pertama iaitu kaedah Baarda, Pope (Tau), Danish dan s:E (Robustified Least Square Estimation) (Halim, 1996). Selisih kasar dikesan melalui pengujian ?erpiawai yang dihitung dengan melibatkan elemen pepenjuru daripada kovarian reja sahaja.

Dua kaedah lain yang diperkenalkan oleh Ethrog (1991) dan Zhang (1987) terlebih dahulu menghitung selisih sebenar dengan mengambilkira korelasi reja.

2.0 KONSEP ASAS PELARASAN L_1

Setiap cerapan l_i dan reja v_i boleh dinyatakan sebagai fungsi terbezakan daripada vektor parameter x_j :

$$l_i + v_i = f_i(x_j)_1 \; ; \; i = 1, 2, .., m \; ; \; j = 1, 2, .., n \; ; \; m > n . \qquad (1)$$

Pelinearan persamaan (1) dengan siri Taylor peringkat pertama memberikan :

$$l_i + v_i = f_i(x^0) + \partial f_i(x)/\partial x_j \big|_{x=x0} \, . \, (x - x_0) \qquad (2)$$

Dalam bentuk matrik ditulis sebagai :

$$A \, . \, \Delta = b + v , \qquad (3)$$

dimana $A = \partial f_i(x)/\partial x_j \big|_{x=x0}$: matrik rekabentuk (m , n)

$\Delta = x - x_0$. : matrik parameter (n , 1)

$b = l - f(x^0)$. : matrik tikaian (m , 1)

v : matrik reja (m , 1).

Secara definisinya, pelarasan dengan L_1 ialah pelarasan parameter Δ dengan memini-makan jumlah mutlak reja cerapan :

$$\sum_{i=1}^{m} |v_i| \quad minima. \qquad (4)$$

sekaligus memenuhi hubungan fungsian parameter-reja (3):

Penerbitan rumusan L_1 diberikan dalam Ralston (1978) dan Marshal & Bethel (1996b). Penyelesaian pelarasan L_1 melibatkan program linear yang memerlukan semua pembolehubah (reja **v** dan parameter Δ) bukan negatif, sedangkan reja cerapan dan parameter sebenar boleh bernilai positif atau negatif. Oleh sebab itu, ekspressi pembolehubah **v** dan Δ perlu disesuaikan menjadi:

$$\mathbf{v}_i = \mathbf{v}^+_i - \mathbf{v}^-_i \, ,$$
$$\Delta_j = \Delta^+_j - \Delta^-_j$$
$$\mathbf{v}^+_i , \mathbf{v}^-_i , \Delta^+_j , \Delta^-_j \quad \geq 0 \text{ untuk } i = 1, 2, ..., m ; \; j = 1, 2, ..., n$$

Perhatikan pula bahawa $|\mathbf{v}_i| = |\mathbf{v}^+_i - \mathbf{v}^-_i| = \mathbf{v}^+_i + \mathbf{v}_i^-$ jika dan hanya jika $\mathbf{v}_i = \mathbf{v}^+_i$ dan $\mathbf{v}^-_i = 0$; atau $v_i = -v_i^-$ dan $v^+_i = 0$. Selanjutnya persoalan L_1 boleh ditulis dalam bentuk program linear:

$$\text{Minimakan:} \quad \sum_{i=1}^{m}(v^+_i + v_i^-), \qquad (5)$$

$$\text{sesesuai dengan} \quad A \, . \, \Delta^+ - A \, . \, \Delta^- - v^+ + v^- = b \qquad (6)$$

Rumusan (5) dan (6) adalah bentuk daripada program linear. Ianya boleh diselesaikan dengan kaedah simplek. Memanfaatkan struktur khas daripada sistem persamaan di atas, Barrodale & Roberts (1973) berjaya mengembangkan satu algoritma yang lebih cekap dan boleh menjimatkan memori komputer. Perisian bagi mendapatkan penyelesaian program linear dengan kaedah simplek berasaskan algoritma tersebut di atas yang ditulis dalam kod Fortran diberikan oleh Branham (1990). Perisian tersebut boleh terus digunapakai dengan input matrik rekabentuk A dan matrik b, sesuai dengan rumusan (5) dan (6), namun ianya belum lagi mengambil kira pemberat cerapan.

Dalam program linear, v^+_i dan v_i^- disebut sebagai pembolehubah 'slack' (slack variable). Dengan wujudnya pembolehubah slack ini satu penyelesaian awal dapat diperolehi dengan cara menetapkan salah satunya v^+_i atau v_i^- sebagai pembolehubah basis. Pembolehubah basis adalah pembolehubah yang nilainya dihitung dengan menganggap pembolehubah non-basis sebagai memiliki nilai sifar. Pada peringkat awal, penyelesaian daripada suatu program linear seperti rumusan (5) dan (6) adalah $v^-_i = b_i$ (untuk $b > 0$) atau $v_i^- = - b_i$ (untuk $b<0$). Proses selanjutnya adalah melakukan saling-tukar ('interchange') basis sehingga diperolehi jumlah mutlak reja minima. Satu algoritma kaedah simplek yang khas dikembangkan oleh Barrodale & Roberts (1974) untuk pelarasan L_1 memudahkan langkah saling-tukar pembolehubah basis ini. Jika matrik A mempunyai takat ('rank') k, maka selalunya akan diperolehi paling sedikit k parameter Δ^+ atau Δ^- sebagai pembolehubah basis dan paling sedikit k pembolehubah v^+_i atau v_i^- akan bernilai sifar.

Berdasarkan penjelasan di atas boleh disimpulkan bahawa penyelesaian pelarasan L_1 diperolehi menerusi pemilihan satu subset cerapan tak berlebih yang memberikan jumlah mutlak reja minima dalam mana subset cerapan itu akan mempunyai reja sifar. Penjelasan kaedah simplek dapat dilihat dalam banyak rujukan, contohnya Noble (1969).

2.1 Pemberat

Apabila cerapan memiliki kejituan yang berbeza, maka konsep pemberat perlu digu-nakan. Pemberat boleh ditakrifkan sebagai darjah kepentingan satu cerapan terhadap cerapan yang lain dalam kumpulannya (Marshall & Bethel, 1996a). Dalam L_1, pemberat cerapan tak berkorelasi ditakrifkan sebagai berikut (Marshall & Bethel, 1996a; Fuchs, 1982; Harvey, 1991):

$$P = \begin{bmatrix} 1/s_1 & 0 & 0 & . & 0 \\ 0 & 1/s_2 & . & & 0 \\ & & . & . & \\ . & & & & 0 \\ 0 & . & & 0 & 1/s_n \end{bmatrix} \tag{7}$$

s_i adalah sisihan piawai cerapan dan $1/s_i = p_i$.

Dalam pelarasan L_1, pemberat dilibatkan dalam bentuk:

$$\text{Minimakan:} \quad \sum_{i=1}^{m}(v^+_i/s_i + v_i^-/s_i), \tag{8}$$

berasaskan persamaan cerapan-parameter (6) :

$$A \cdot \Delta^+ - A \cdot \Delta^- - v^+ + v^- = b \,.$$

Untuk membolehkan penggunaan terus algoritma Simplek yang diubahsuai oleh Barrodale & Robert (1973), maka persamaan cerapan-parameter di atas diubah menjadi persamaan yang setara dengan mendarabkan sisi kiri dan kanan dengan matrik pemberat P (Fuchs, 1982):

$$P. A \cdot \Delta^+ - P. A \cdot \Delta^- - P.v^+ + P. v^- = P \cdot b \tag{9}$$

Disebabkan P adalah matrik pepenjuru dengan elemen p_i, maka satu persamaan cerapan di atas berikut boleh disusun semula menjadi (Fuchs, 1982):

$$A' . \Delta^+ - A' . \Delta^- - v'^+ + v'^- = b' \qquad (10)$$

Dimana $a'_{ij} = a_{ij} . p_i$, $b'_i = b_i . p_i$, $v'^+ = v^+ . p_i$ dan $v'^+ = v^+ . p_i$.

Perisian LINDO (Linear Integer and Discrete Optimizer) boleh juga boleh digunapakai bagi menyelesaikan pelarasan L_1. Menerusi LINDO, data input untuk program linear terus dituliskan seperti rumusan aljabar (5) dan (6) atau (8) dan (9) (Winston, 1987)

2.2 Pengesanan Selisih Kasar

Pelarasan L_2 sangat sensitif terhadap wujudnya selisih kasar, iaitu selisih kasar yang wujud akan mengkontaminasi cerapan lain sehingga pengesanan selisih kasar menjadi lebih sukar (seksyen 1.0). Lebih spesifik lagi dikatakan oleh Chen, Kavouras & Chrzanowski (1987) bahawa kesukaran akan bertambah dengan besarnya korelasi antara reja, bilangan selisih kasar dan rendahnya bilangan lebihpada (redundancy number). Pada amnya, kesukaran utama dalam pengesanan selisih kasar disebabkan oleh kesukaran memperolehi reja cerapan yang hampir kepada keadaan selisih sebenar.

Satu prosidur pengesanan selisih kasar berasaskan L_1 telah diperkenalkan oleh Gao , Krakiwsky & Czompo (1992). Penjelasan ringkas prosidur diberikan di bawah ini (lihat Rajah 1).

Untuk satu model cerapan-parameter (**A, b, C**$_l$) di mana C_l matrik kovarian cerapan, A matrik rekabentuk takat penuh (full rank), setelah pelarasan L_1 dilakukan, satu pemetakan matrik boleh dibentuk sebagai:

$$\begin{bmatrix} \mathbf{A}_1 \\ (n , n) \\ \mathbf{A}_2 \\ (m\text{-}n , n) \end{bmatrix} \underset{(n,1)}{\Delta} = \begin{bmatrix} \mathbf{b}_1 \\ (n , 1) \\ \mathbf{b}_2 \\ (m\text{-}n , 1) \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ (n , 1) \\ \mathbf{v}_2 \\ (m\text{-}n , 1) \end{bmatrix} \qquad (11)$$

v_1 ialah reja bernilai sifar manakala v_2 merujuk kepada cerapan yang mengandungi reja tidak sifar, termasuk selisih kasar bila wujud. Selanjutnya boleh ditulis hubungan di bawah ini :

$$\Delta = (A_1)^{-1} . b_1$$
$$v_2 = b_2 - A_2 . \Delta = b_2 - A_2 (A_1)^{-1} b_2$$

Bila dianggap cerapan tidak berkorelasi, kovarian reja adalah:

$$C_{v2} = C_{l2} + A_2 (A_1)^{-1} . C_{l1} (A_2 (A_1)^{-1})^T \qquad (12)$$

Selanjutnya ujian global dilakukan dengan rumusan:

Hipotesis nol statistik ujian global ialah :

$$T' = (v_2)^T . C_{v2} . v_2 \,\Big|\, Ho \sim \chi^2 (0): \qquad (13)$$

Untuk tahap keertian α, pada darjah kebebasan r, ujian global satu sisi lulus apabila :

$$T' < \chi^2_{r, \alpha} \qquad (14)$$

Apabila ujian global gagal, maka ujian lokal untuk mengesan selisih kasar dilakukan se-terusnya. Statistik ujian ialah:

$$w' = v_2 / (C_{v2ii})^{1/2}$$

Sebagai hipotesis nol H_o, dikatakan w' mengikut taburan normal piawai:

$$w' = v_2 / (C_{v2ii})^{1/2} \quad H_o \sim N(0, 1) \tag{15}$$

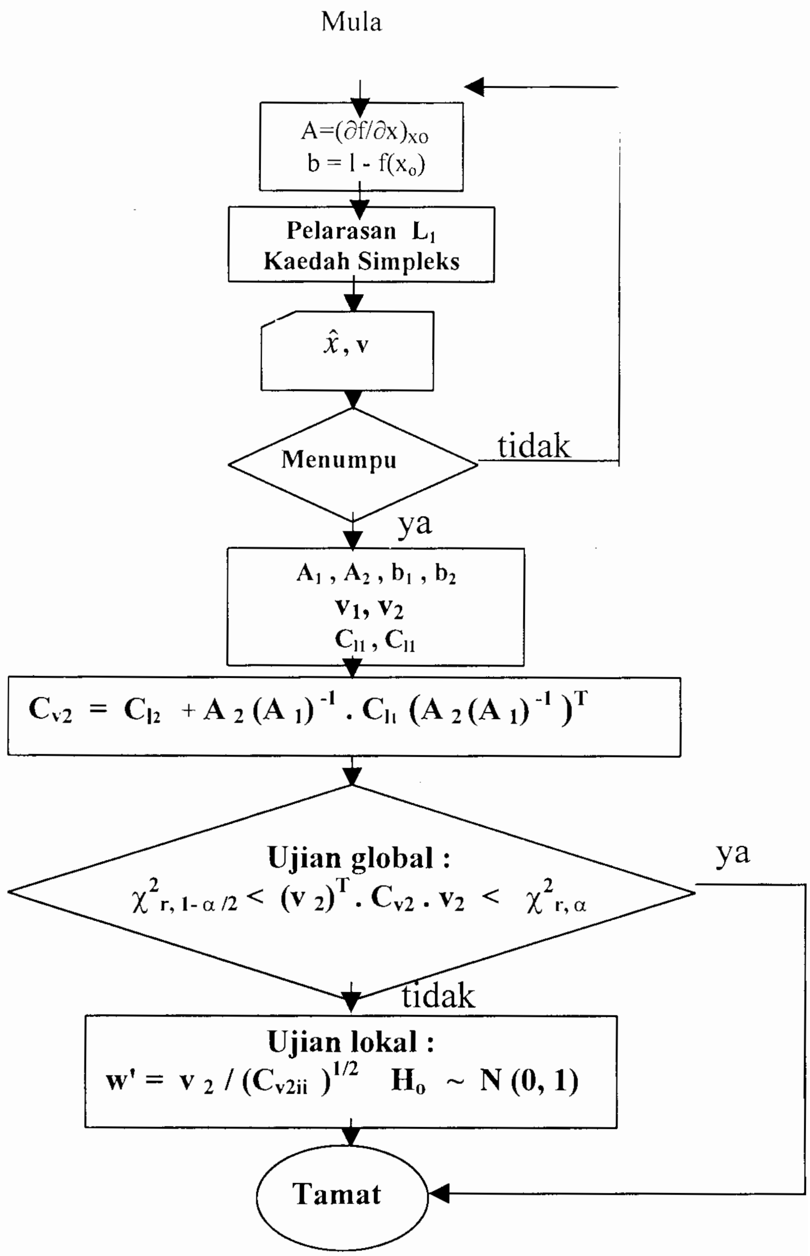
Prosidur pengesanan selisih kasar di atas mengandungi dua anggapan: reja cerapan yang dihitung menerusi pelarasan L_1 mengandungi anggapan bahawa selisih mengikut taburan tidak normal; sedangkan penerbitan statistik ujian dilakukan dengan menganggap ianya mengikut taburan normal. Marshal & Bethel (1996a) menyatakan bahawa penggunaan andaian model stokastik pada pelarasan L_1 menghalangi penggunaan ujian statistik konvensional, termasuk juga yang diajukan oleh Gao, Krakiwsky & Czompo (1992). Untuk mendapatkan analisis statistik terhadap reja hasil daripada pelarasan L_1, satu simulasi Monte Carlo bagi mendapatkan ciri-ciri daripada taburan kebarangkalian reja perlu dilakukan (Fuch, 1982; Marshal & Bethel, 1996a). Namun melakukan Simulasi Monte Carlo memakan masa (Barradole & Roberts, 1973) dan tidak praktikal, terutamanya apabila ukuran daripada matrik rekabentuknya besar, dan jumlah cerapan jauh lebih besar daripada jumlah parameter.

Model empirikal daripada reja hasil pelarasan L_1 boleh didapati dengan menggunakan simulasi kaedah Monte Carlo. Dalam simulasi Monte Carlo, setiap gangguan rawak diberikan kepada vektor b_i. dan reja dihitung semula menerusi pelarasan L_1. Fuch (1982) dan Marshall & Bethel (1996a) melakukan simulasi dengan gangguan rawak mengikut taburan normal, masing-masingnya pada satu jaringan 'direction-distance' dan trilaterasi. Marshall & Bethel (1996a) melakukan sebanyak 100,000 simulasi untuk mendapatkan taburan reja empirikal mengikut kepada persamaan eksponensial : f(x) = a * (e** (b * (v**c))). Dari simulasi Monte Carlo diperolehi satu taburan empirikal sampel reja. Berdasarkan taburan empirikal ini ujian statistik dilakukan, termasuk ujian wujudnya selisih kasar

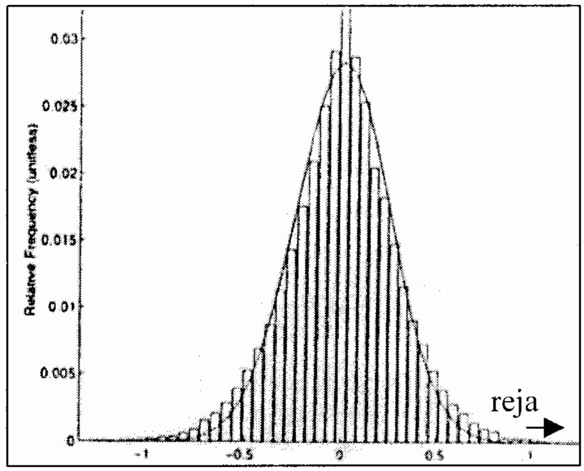
Hasil daripada simulasi Monte Carlo yang dilakukan oleh Marshall & Bethel (1996a) memperlihatkan ketidak-normalan taburan reja (Rajah 2). Taburan empirikal mempunyai puncak yang lebih tinggi (sebahagian dipotong) dan 'tail' (ekor) yang lebih lebar dan tebal berbanding dengan lengkung taburan normal yang dihitung daripada data yang sama. Sama ada taburan empirikal dan normal mempunyai paksi simetri pada puncak, yaitu pada reja sifar. Kebarangkalian pada taburan normal lebih cepat menuju 100 peratus berbanding taburan empirikal.

Tetapi dapat dilihat bahawa taburan empirikal akan lebih menghampiri normal apabila reja bernilai sifar (v_1 pada penyelesaian L_1 - bahagian 'peak' pada taburan empirikal) tidak diambil kira dalam melakukan 'curve fitting' persamaan eksponesial di atas.

Ini bermakna bahawa andaian yang digunakan untuk menerbitkan statistik ujian yang digunakan oleh Gao, Krakiwsky & Czompo (1992) cukup sesuai dengan kenyataan empirikal yang ditemukan oleh Marshal & Bethel (1996a) tersebut. Menggunakan prosidur seperti yang diperkenal di atas, satu ujian pengesanan selisih kasar terhadap data daripada Andrew (1974) telahpun berjaya mengesan selisih kasar yang wujud dan kesimpulannya sama dengan apa yang dilakukan oleh Andrew tersebut (Gao, Krakiwsky & Czompo, 1992). Namun keberkesanan prosidur tersebut terhadap sebarang selisih kasar mengikut beberapa kondisi seperti bilangan lebihpada, korelasi antara reja dan jumlah daripada selisih kasar itu sendiri belum lagi dikenal pasti. Dalam kajian ini telah dilakukan terhadap data daripada Zhang (1987) dan ianya memberikan keberkesanan yang amat baik. Selanjutnya boleh dilihat pada contoh numerik di bawah ini.



Rajah 1 : Cartalir Pengesanan Selisih Kasar Berbasiskan L_1



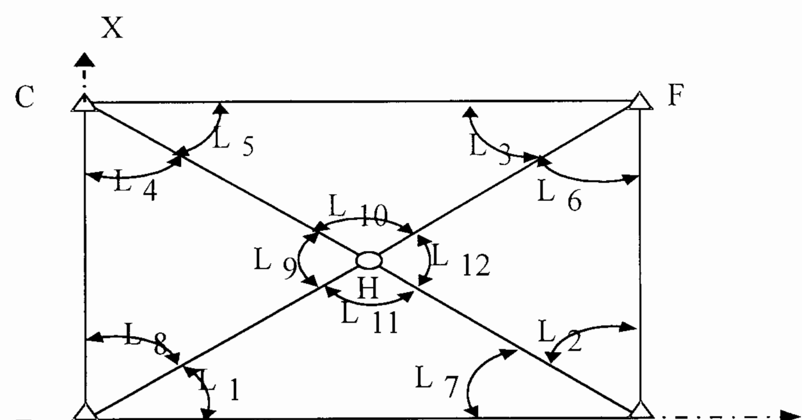
Rajah 2:
Taburan kebarangkalian empirikal reja : hasil daripada Simulasi Monte Carlo (Marshall dan Bethel, 1996)

3. CONTOH NUMERIK

Daripada data Zhang (1987) seperti pada Rajah 3 diberikan sisihan piawai cerapan adalah sama (5") dan koordinat anggaran titik H : X_0=50.00 ; Y_0=50.00.

$$A^T = \begin{bmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 0 & 2 & -2 & 0 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -2 & 0 & 0 & 2 \end{bmatrix}$$

$$b^T = \begin{bmatrix} -5'' & -2'' & 4'' & 20'' & 0'' & -30'' & 5'' & 3'' & -5'' & -2'' & 1'' & 25'' \end{bmatrix}$$



Rajah 3: Data daripada Zhang (1987)

Pengesanan selisih kasar berasaskan L_1 pada tahap keertian 0.1% (nilai kritikal w'= 3.1) memberikan kesimpulan yang sama dengan Zhang (1987), yaitu cerapan 4, 6 dan 12 (bertanda *) mengandungi kasar manakala pada tahap keertian 5% (nilai kritikal w' = 1.65) reja yang dikesan sebagai selisih kasar adalah 1, 4, 6 dan 12 (bertanda * dan **). Penyelesaian dengan L_1 memberikan satu subset cerapan tak lebih pada yang, yaitu cerapan 3 dan 11. Reja daripada subset cerapan ini bernilai sifar. Di bawah ini adalah jadual reja petakan 2 (v_2).

i	V2 (second)	W'
1**	-20.0626	2.837
2	2.0626	0.238
4*	33.0024	3.811
5	6.1879	0.714
6*	-51.5662	7.292
7	4.1253	0.476
8	-2.0626	0.292
9	4.1253	0.337
10	0.0000	0.000
12*	49.9161	7.059

Jadual 1: Reja petakan 2 (v_2)

3.0 KESIMPULAN DAN CADANGAN

Dari contoh numerik yang diberikan di atas menunjukkan bahawa prosidur pengesanan selisih kasar berasaskan kepada pelarasan L_1 memberikan keberkesanan yang baik. Satu ujian terhadap data 'multilinear regression' yang dilakukan oleh Gao, Krakiwsky & Czompo (1992) berasaskan kepada pelarasan L_1 memberikan pula kesimpulan yang sama, iaitu selisih kasar dapat dikesan dengan baik.

Ujian terhadap sebarang selisih kasar bergantung kepada faktor berikut: korelasi antara reja, bilangan lebihpada dan bilangan selisih kasar yang wujud perlu dilakukan untuk mendapatkan gambaran sebenar daripada keberkesanannya. Satu perisian khas bagi mengesan selisih kasar berdasarkan pelarasan L_1 sedang dibangunkan.

5.0 PENGHARGAAN

Kajian ini adalah sebahagian daripada projek penyelidikan IRPA melalui vot 72070.

RUJUKAN

Barrodale. I & Roberts, F.D.K (1973) **An Improved Algorithm for Discrete L_1 Linear Approximation**, SIAM J. Numer. Anal..Vol. 10. No.5.

Branham. R.L. Jr. (1990) **Scientific Data Analysis: An Introduction to Overdetermined System**, Springer-Verlag Newyork Inc.. 237p.

Chen. Y.Q., Kavouras, M., & Chrzanowski, A. (1987) **A Strategy for Detection of Outlying Osevations in Measuremcnts of High Precision,** The Canadian Surveyor, Vol 41, No.4, pp 529-540.

Ethrog, U (1991) **Statistical Test of Significance for Testing Outlying Observation**, Survey Riview, 31, 240, pp. 62-69

Fuchs, H. (1982) **Contribution to Adjustment by Minimizing the Sum of Absolute Residuals**, Manuscripta Geodaetica. Vol. 7 pp. 151-207.

Gao Y., Krakiwsky E. J. , & Czompo J. (1992). **Robust Testing Procedure for Detection of Multiple Blunders**. Journal of Surveying Engineering 118(1), pp. 11-23.

Halim Setan (1995) **Functional and Stochastic Model for Geometrical Detection of Spatial Defor-mation in Engineering: A Practical Approach**. PhD Thesis, ESRC, Department of Civil Engineering, City University, London.

Halim Setan (1996). **A Practical Strategy For Detecting Multiple Gross Errors**, Buletin ukur, Jilid 7, No. 1.

Harvey, B.R. (1991) **Survey Network Adjustment by The L_1 Method**. AJGPS No. 59, pp. 39-52.

Marshall, J. & Bethel, J. (1996a) **Analysis Residuals From L_1 Norm Estimation**, International Archives of Photogrametry and Remote Sensing. Vol. XXXI, Part B5, Viena.

Marshall J. & Bethel J. (1996b) **Basic Concepts of L_1 Norm Minimization For Surveying Applications,** Journal of Surveying Engineering 122(4), pp. 168-179.

Noble, B. (1969) **Applied Linear Algebra,** Precentice-Hall, New Jersey.

Ralston, A. (1978) **International series in Pure and Applied Mathematics : A First Course in Numerical Analysis**, Mc-Graw-Hill Book Company.

Winston, W. L. (1987) **Operation Research : Application and Algorthms**, Duxbury Press Boston.

Zhang Bingcai. (1987) **A New Method of Data Snooping** , AJGPS, No. 46 & 47, pp. 103-122.